

One Platform for Data Observability Across Global Data Divisions



CLIENT

Global biopharmaceutical company operating across multiple divisions.

THE CHALLENGE

The client needed a unified way to manage a fragmented data ecosystem, control rising operational costs, and protect data integrity across its divisions.

- No centralized view into the health, quality, and cost of analytical assets
- Siloed monitoring development was estimated at three times the cost of a single unified platform
- High risk of misinformed business decisions due to the absence of proactive anomaly detection
- Hundreds of project owners had no automated way to track their own resource use and budget consumption



SOLUTION

UD4D partnered with a client to architect a modular, cloud-native platform on AWS to centralize metadata and automate observability across the data estate.

- Metadata integration across global Databricks workspaces and GPTeal (generative AI program), including definition and ingestion of critical metadata
- Validation of Redshift data models through review and optimization, ensuring a high-performance foundation for analytics in Power BI and ThoughtSpot
- Federated data access through a Trino layer over processed observability data, enabling cross-source querying
- Specialized visualization and reporting via 4 Grafana dashboards for Databricks and daily summary email reports for Redshift usage analytics
- Secure access framework using AWS IAM and SSO for granular, workspace-specific data isolation

BUSINESS IMPACT

50% **faster issue resolution:** Time to troubleshoot data and pipeline issues was cut in half.

3 **proactive FinOps alerts:** Automated alerts flag cost spikes above 20% and hardware utilization anomalies above 80% or below 10%.

3x **lower development cost:** Consolidating monitoring into a single platform avoided the estimated triple cost of siloed, per-division tooling.

2 **minute dashboard loads:** Responsive dashboards load in under 2 minutes and filters refresh in under 1 minute.

TECHNOLOGY STACK

AWS (S3, Glue, Redshift, IAM) | Databricks | Trino | Apache Airflow Grafana
| Collibra | Power BI | ThoughtSpot